Patient income and health innovation

Javad Moradpour, Aidan Hollis.

Department of Economics, University of Calgary

October 2019

Abstract

This paper proposes and tests a model of investment into health innovation across therapeutic categories. It explicitly analyzes the relationship between the number of clinical trials in a disease area, the health losses from that disease, and the average income of people suffering from it. Average patient income is strongly predictive of the number of clinical trials, whether funded by industry or not. We are able to precisely estimate the extent of this income effect on the number of trials, and to identify both (a) the specific diseases that appear to be underfunded relative to their harm to human health and (b) the amount of additional funding required to bring innovation investment up to the present average.

Introduction

Given that most pharmaceutical research is directed by for-profit companies, it is not surprisingly disproportionately focused on the needs of people in rich countries. Globally, the number of deaths from malaria and diabetes is similar; but there are about 10 times as many clinical trials for diabetes, which affects much wealthier patients on average. How important is patient income in determining which diseases attract drug research? How much investment is needed to adequately fund innovation into therapies for diseases mainly suffered by people in poor countries? Is there a way to systematically identify what diseases should receive more investment?

These questions have not been adequately addressed. In the global health literature, the "10/90 Gap", expressing the idea that only 10% of health research addresses the causes of 90% of the global burden of disease, has become conventional wisdom. (Commission on Health Research for Development, 1990; Labonte & Spiegel, 2003; Lown & Banerjee, 2006; Vidyasagar, 2006) It is, however, far from accurate. The premise is that research focused on the diseases of low- and middle-income countries is proportionately very small in relation to excess mortality in those countries. What makes this claim problematic is that people in all countries suffer from global diseases, such as cancers and cardiovascular disease, which create much of the burden of disease everywhere. Research on global diseases is thus relevant to the needs of all.

Understanding how health research is directed is important for policy. For example, because of perceived under-financing of research related to "neglected tropical diseases," the US created a special incentive for drugs treating these diseases.(Ridley, Grabowski, & Moe, 2006; US Food and Drug Administration, 2016) A few years later, there was a contrary focus on non-communicable diseases, with a UN General Assembly Resolution in 2012 encouraging the development of new medicines to prevent and treat them.(UN General Assembly, 2012) It would be helpful to have a systematic way of identifying which diseases are most underfunded and by how much.

We develop a model of investment in clinical trials, provide evidence on how well the distribution of clinical trials matches the burden of disease globally, and show how the incomes of people suffering a disease relate to the number of clinical trials for it. Both results are novel. Numerous previous studies have examined the relationship between innovation and the burden of disease, generally finding a small positive, or no, correlation;

but no studies to our knowledge have directly examined the effect of income on the number of clinical trials across therapeutic classes.

Prior literature on the relationship between health losses and innovation

Several studies have sought to identify the relationship between market size and the development of new pharmaceuticals, but while this is connected to our research, it differs in important ways. (Acemoglu & Linn, 2004; Dubois, de Mouzon, Scott-Morton, & Seabright, 2015; Kyle & McGahan, 2011) Market size is some composite of health needs and income, which we separate in our analysis. In addition, given data challenges, these prior studies have widely varying estimates of the responsiveness of drug development to market size.

Existing studies on the burden of disease and the amount of research or new drugs developed have mixed results. Viergever *et al* find "little correlation" between the burden of disease and the number of trials. (Viergever, Terry, & Karam, 2013) Martino *et al* report no relationship between burden of disease and innovation.(Martino, Ward, Packer, Simpson, & Stevens, 2012) Lichtenberg, in contrast, finds the elasticity of the number of chemotherapy regimens and MEDLINE drug citations with respect to global cancer incidence to 0.53 and 0.60 respectively. (Lichtenberg, 2007)

Barrenho *et al* consider the burden of disease and market size, where market size for disease *i* is constructed as $\sum_j DALYs_{ji} \times GDP_j$, and *j* indexes countries, and explore the relationships with the number of new drugs.(Barrenho, Miraldo, & Smith, 2019) As with our paper, they rely on the Global Burden of Disease data and GDP information, but do not separate out income from the burden of disease. Since, as we find, the responsiveness of

investment differs between income and disease burden, it is helpful to distinguish them. Barrenho *et al* analyze individual disease areas to identify whether innovation matches the disease burden. Their approach is to use concentration curves and indices rather than a regression framework.

In summary, the existing empirical evidence on the relationship between clinical research and clinical importance of diseases is unclear; and there is little direct evidence on how patient income relates to investment in innovation. This paper presents novel evidence on both points within the context of an illustrative theoretical model. In addition, the calculation of "average patient income", while straightforward, is both new and useful. Finally, the empirical strategy used allows one to identify diseases that appear to be under-researched because of low average patient income.

Model

We develop a simple model of differentiated goods in a market with monopolistic competition, based on Perloff & Salop (1985). This model is intended to provide intuition rather than to be a complete description of pharmaceutical innovation. Each disease $z \in 1,2, \dots, Z$ has potential patients subscripted $i = 1, \dots, I_z$, and treatments subscripted $t = 1, \dots, T_z$. For each disease, we denote the maximum health losses that could be caused by the disease to an individual by m_z, and the average income of patients by y_z. Patient preferences over treatments are such that the surplus obtained from treatment t is given by $s_{itz} = \theta_{itz} - \frac{p_{tz}}{\phi_z}$. The therapeutic "value" of treatment t to patient *i* with disease *z* is represented by $\theta_{itz} \sim U[0, m_z]$, so treatments vary in their average value across diseases, and in their specific value to each patient with a given disease. The price of treatment t is

denoted p_{tz}, and ϕ_z represents anything specific to the disease that affects willingness to pay, excluding the therapeutic value of the treatment. We assume that $\max_t s_{itz} \geq 0$ for each patient, and that no patient wants to buy more than one treatment per disease, so that every patient will want to buy exactly one treatment for any disease affecting them.

Given the prices and therapeutic value of each treatment, patients will choose the treatment for which their surplus s_{itz} is maximized. If $s_{itz} \geq s_{i\hat{t}z}$ (where $\hat{t} \neq t$) for consumer i, then $\theta_{i\hat{t}z} \leq \frac{p_{\hat{t}z} - p_{tz}}{\phi_z} + \theta_{itz}$. Thus, given θ_{itz}, the probability that $s_{itz} \geq s_{i\hat{t}z}$ is $Pr(s_{itz} \geq s_{i\hat{t}z}) = \frac{p_{\hat{t}z} - p_{tz}}{m_z \phi_z} + \frac{\theta_{itz}}{m_z}$, which is the cumulative distribution.

Since the θ_{itz} are distributed independently, the expected proportion of consumers who purchase treatment t is given by

$$Pr\left(s_{itz} \geq \max_{\hat{t} \neq t} s_{j\hat{t}z}\right) = \int_0^{m_z} \prod\nolimits_{j \neq i} \left(\frac{p_{\hat{t}z} - p_{tz}}{m_z \phi_z} + \frac{\theta_{tz}}{m_z}\right) \frac{1}{m_z} d\theta_{tz}$$

The expected number of units sold of each treatment will be equal to the proportion of consumers who prefer that treatment times the number of patients affected by that disease I_z:

$$Q_{tz}(p_{1z}, .., p_{t_z z}) = Pr\left(s_{itz} \geq \max_{\hat{t} \neq t} s_{i\hat{t}z}\right) I_z$$

Assuming that each product has a constant marginal cost c and research and development costs of K, the expected profits from each therapy are given by

$$\pi_{tz} = (p_{tz} - c) Q_{tz}(p_{1z}, .., p_{t_z z}) - K \qquad (1)$$

.

We assume that each risk-neutral firm maximizes expected profits, and we apply a Bertrand-Nash framework for these differentiated goods. We also assume that firms do not

have multiple products competing against each other in a given therapeutic class. This implies that

$$p_{tz} = c - \frac{Q_{tz}(p_{1z}, .., p_{t_z z})}{\partial Q_{tz}/\partial p_{tz}} \tag{2}$$

Given that θ_{itz} has the same distribution for all patients with a given disease, we will also assume a single-price equilibrium. We then write $p_{tz} = p_z$ for all $t = 1, \dots, T_z$. The symmetric industry equilibrium has

$$Q_{tz}(p_z, .., p_{tz}, \dots, p_z) = T_z \int_0^{m_z} \left(\frac{p_z - p_{tz}}{m_z \phi_z} + \frac{\theta_{itz}}{m_z}\right)^{T_z - 1} \frac{1}{m_z} d\theta_{tz}$$

and the slope of demand for treatment *t* is given by

$$\frac{\partial Q_{tz}}{\partial p_{tz}} = -\frac{T_z - 1}{m_z \phi_z} I_z \int_0^{m_z} \left(\frac{p_z - p_{tz}}{m_z \phi_z} + \frac{\theta_{tz}}{m_z}\right)^{T_z - 2} \frac{1}{m_z} d\theta_{tz}$$

After setting $p_{tz} = p_z$, we obtain

$$Q_z(p_z) = I_z \int_0^{m_z} \left(\frac{\theta_z}{m_z}\right)^{T_z - 1} \frac{1}{m_z} d\theta_z = \frac{I_z}{T_z} \tag{3}$$

so that in expectation patients are equally distributed among the different treatments. The slope of demand can then be written as

$$\frac{\partial Q_{tz}}{\partial p_{tz}} = -\frac{T_z - 1}{{m_z}^{T_z} \phi_z} I_z \int_0^{m_z} (\theta_{tz})^{T_z - 2} \, d\theta_{tz} = -\frac{I_z}{m_z \phi_z} \tag{4}$$

Substituting (3) and (4) into (2), we obtain a equilibrium price:

$$p_z = c + \frac{m_z \phi_z}{T_z}$$

Expected profits are then

$$\pi_z = \frac{m_z \phi_z I_z}{{T_z}^2} - K$$

Assuming monopolistic competition, expected profits will be driven to zero, so that the equilibrium number of treatments for disease z is given by

$$T_z^* = \sqrt{\frac{m_z \phi_z I_z}{K}} \tag{5}$$

i.e. the number of treatments T_z for disease z will be positively related to the health burden per patient m_z, the willingness to pay measure ϕ_z, and the number of individual patients affected I_z, and negatively related to the fixed cost of developing a treatment. This model sacrifices completeness for tractability in order to suggest a partial framework for thinking about how firms choose which disease areas to target. A more complete model of firms' decisions would account for existing therapies, the state of scientific knowledge, the potential for price discrimination across countries, insurance, differences across health systems, and the like. However, this simple model helps to illuminate the dependence of innovation on the key factors of severity, willingness to pay, and the number of patients.

Empirical strategy

We can rewrite (5) as

$$\log(T_z^*) = -\tfrac{1}{2}\log(K) + \tfrac{1}{2}\log(\phi_z) + \tfrac{1}{2}\log(m_z I_z)$$

Recall that ϕ_z describes aspects of the disease that might be related to willingness to pay (separate from the health impact of the disease itself). Thus, suppose that ϕ_z is related to income according the following:

$$\log(\phi_z) = \alpha_0 + \alpha_1 \log(y_z) + \varepsilon$$

Income, of course, doesn't only directly affect willingness to pay; patients who live in higher-income countries tend to have more efficient pharmaceutical distribution systems,

better access to physicians and diagnostics, and more comprehensive health insurance. All of these are related to income and income therefore also acts indirectly on willingness to pay. The relationship between ϕ_z and y_z may reflect all of these effects.

Thus, we can estimate the equation:

$$\log(T_z^*) = \beta_0 + \beta_1 \log(y_z) + \beta_2 \log(m_z I_z) + u$$

in which the model predicts that $\beta_2 = 0.5$. This model requires only that we know about disease-level data, specifically the income of patients y_z and the total burden of disease $m_z I_z$ that can be addressed by new treatments.

Our empirical strategy is based on the assumption that drugs currently in development do not immediately change the burden of disease or the average patient income meaningfully; this assumption removes concerns of endogeneity. There are several reasons for making this assumption. First, most clinical trials are failures, in the sense that they do not result in an approved product. Second, even after a successful trial, it typically takes years before products are approved in most jurisdictions. Third, for most products, commercial take up is slow and peak sales are not achieved for several years. Finally, very few approved products have a significant effect on the global burden of disease even within their disease area. (A significant exception is sofosbuvir, which however only came to market in 2013 and therefore had very little impact on the overall burden of disease even in hepatitis, given the way that we constructed our data as described below.)

Data

We obtained data on the burden of disease for 94 different diseases and conditions (hereafter, "diseases") from the Global Burden of Disease Study and listed in Appendix 1. We focused on DALYs (disability-adjusted life-years lost) which is a summary measure of

health loss (*i.e.* $m_z I_z$), and use the average of the estimates for 2005, 2010, and 2015 (Murray et al., 2012). We limited diseases to those for which there is a plausible pharmacological intervention (excluding, for example, deaths by traffic accident). We excluded residual categories, such as "Other infectious diseases." We excluded 10 causes for which there is an existing excellent treatment, such as leprosy, since such diseases will typically be the subject of few clinical trials. (See Appendix 2, which also shows that inclusion of these diseases does not meaningfully change our results.) We identified each disease as being a "Neglected Tropical Disease" if it was so categorized by the WHO.(World Health Organization, 2017) We also identified each disease as being infectious or not.

We then generated a measure of "average patient income" (hereafter, API) for each disease. The API is calculated as $API_z = \sum_j HL_{zj} G_j \,/\, \sum_j HL_{zj}$ where HL_{zj} represents the health losses of disease z in country j, and G_j represents the annual per capita GNI of country j, averaged over the years 2006 – 2015, in 2010 US dollars. GNI data were extracted from World Bank.(World Bank, 2019) API thus represents the average GNI per capita, weighted by the disease's health losses in each country. For example, malaria's API is $1,255, because most people infected by malaria live in low-income countries; diabetes, with a more even geographical distribution, has an API of $10,097. Evidently, API is only a proxy for y_z, the income of people who have a given disease; our underlying assumption is that ability or willingness to pay is on average related to national per capita income.

Finally, we used clinical trials for each disease as our measure of T_z, as investment in clinical trials is the required precursor to drug approval. Since 2007, US laws have mandated the registration of certain clinical trials in registered in clinicaltrials.gov, and over 250,000 trials are listed. In addition, leading medical journals, since 2004, have

required registration in a "public trials registry" as a condition for publication.(De Angelis et al., 2004; Weber, Merino, & Loder, 2015) We included only Phase 1, 2, 3, or 4 clinical trials. This typically excludes studies of devices or behavioral interventions. (ClinicalTrials.gov, 2017) We attributed 50% of each clinical trial listed as Phase 1/2 to both Phase 1 and Phase 2; and similarly for Phase 2/Phase 3 studies.

Each disease in the Global Burden of Disease Study is described by one or more ICD-9 or ICD-10 codes. We mapped each ICD code to one or more corresponding MeSH categories, using the Disease Ontology database.(Institute for Genome Sciences, 2019) We then identified all clinical trials for each MeSH code, mapped back to the 94 diseases. Clinical trials may list multiple MeSH codes, and in some cases will therefore link to more than one disease. We used clinical trials with an initiation date between 2006 and 2015. There were 60,253 Phase 1, 2, 3, or 4 clinical trials during this period corresponding to the 94 diseases. Summary data for the number of clinical trials for each disease, DALYs, and API are shown in Table 1. Each clinical trial, on average, linked to 1.43 diseases.

We also identified the sponsor type for each clinical trial. Sponsors are classified as government (NIH or U.S. Federal), industry, or other. "Other" mainly indicates universities, hospitals, foundations, and non-US governments. We group "other" with government in our analysis. Of the 60,253 total trials, 32,743 have at least one industry sponsor. Summary statistics for this data is shown in Table 1 below.

Phase	Freq.	Percent
Early Phase 1	901	1.5
Phase 1	10,722	17.79
Phase 1/Phase 2	4,421	7.34
Phase 2	18,795	31.19
Phase 2/Phase 3	2,060	3.42
Phase 3	12,993	21.56
Phase 4	10,361	17.2
Total	60,253	100

Intervention type	Freq.	Percent
Drug or Biological	51,150	84.89
Other	9,103	15.11
Total	60,253	100

Involvement	Freq.	Percent
Industry	32,743	54.34
Non-Industry	27,510	45.66
Total	60,253	100

Variable	Obs	Mean	Min	Max
Clinical trials	94	641	5	7112
DALYs	94	1.65E+07	170028.7	1.54E+08
API	94	10770.8	447.3	27110.8

Table 1. Summary Statistics

The use of clinical trials as a proxy for investment in innovation is motivated by the fact that clinical trials are critical inputs. New therapies cannot be approved without clinical trials; and trials constitute over half the cost of drug development.(DiMasi, Grabowski, & Hansen, 2016) Since the actual cost of each trial is confidential we use a count of trials. We also prefer data on clinical trials as a measure of investment over data on drug approvals, since the former are fully controlled by investors and the latter are not.

Since clinicaltrials.gov is a US database, we had concerns that it might not be representative with respect to clinical trials for diseases primarily affecting low-income countries. Using a global clinical trials registry, we show in Appendix 3 that for the relevant period, the coverage of clinicaltrials.gov is unrelated to API.

Figure 1 illustrates the importance of patient income. Diseases are ranked by API, and the cumulative share of DALYs from those diseases and the related clinical trials are plotted. If API were irrelevant, the clinical trials distribution would be expected to roughly overlay the DALY distribution. As Figure 1 shows, diseases with an API of under $5,000 make up about 28% of DALYs but less than 10% of trials, suggesting substantial underinvestment into clinical trials for these diseases. In contrast, diseases with API above $15,000 represent less than 10% of DALYs but 28% of trials. Another way of expressing this is that the diseases that cause the lowest-income 90% of DALYs are the subject of only 72% of clinical trials – i.e. there is a "72/90 Gap".

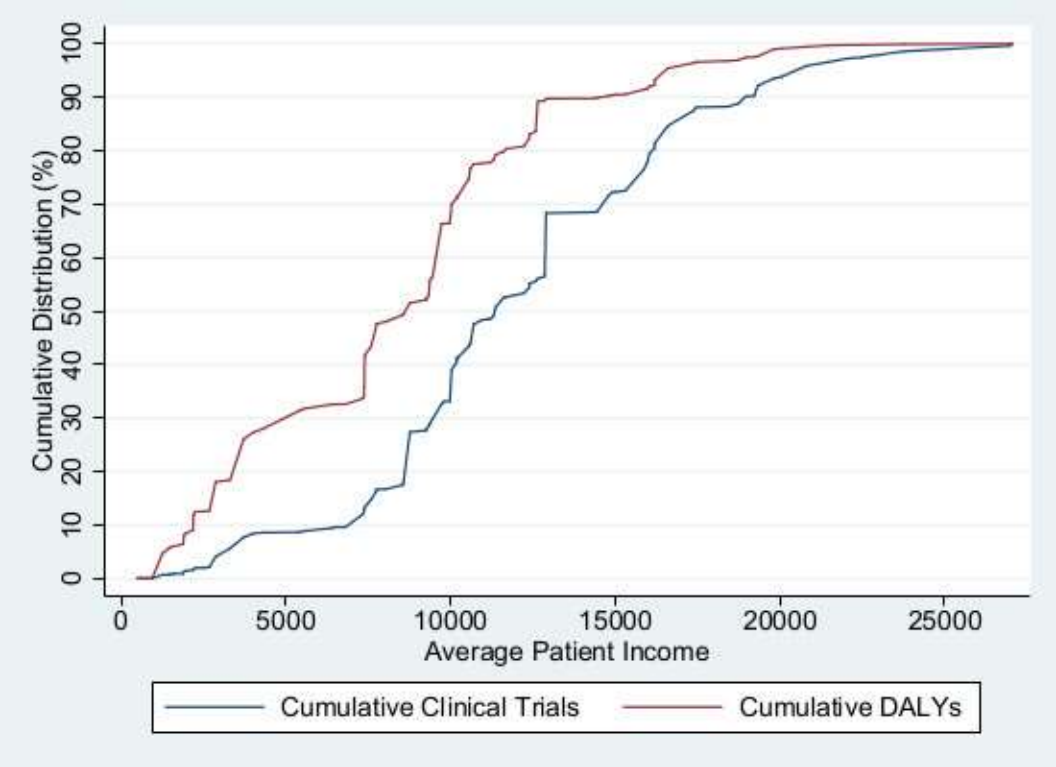

Fig. 1. Cumulative clinical trial and DALY shares ranked by Average Patient Income.

The relationships observed in Figure 1 are explained well by considering Figure 2, which shows how the number of clinical trials relates individually to the number of DALYs in each disease and the patient income. This shows a clear positive relationship between the number of clinical trials and the number of DALYs and API, though we need regression to identify marginal effects.

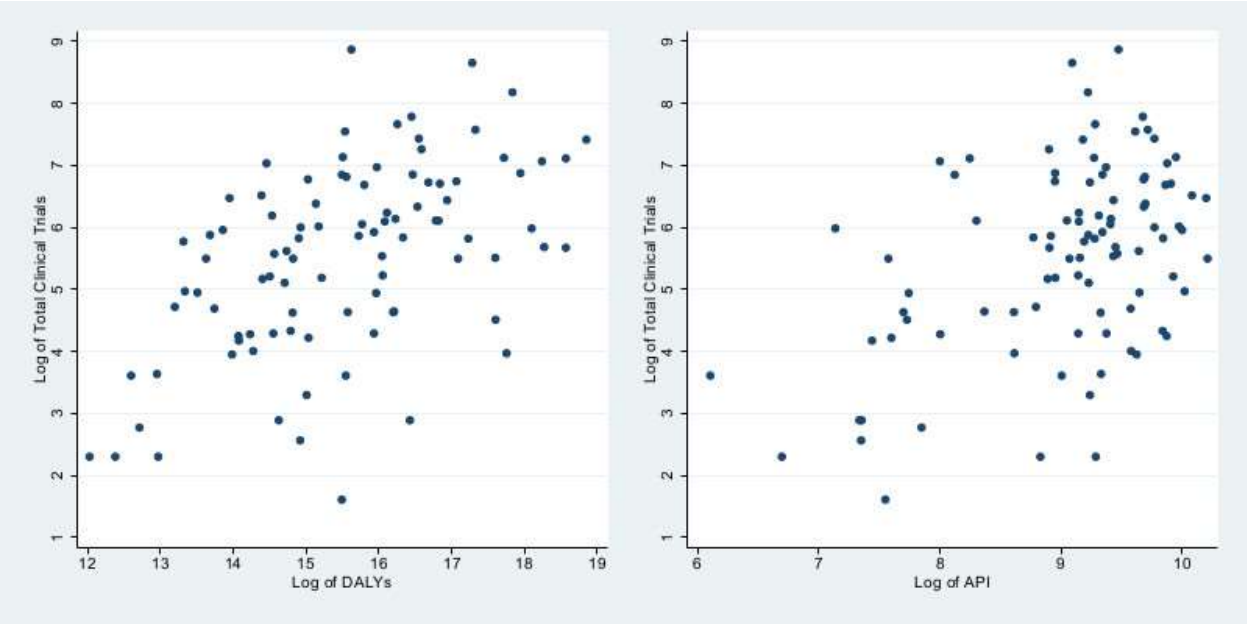

Fig. 2. Clinical trials compared to DALYs and Average Patient Income for each disease.

Results

Given the model developed above, we use a logarithmic transformation of variables to quantify the relationship between the number of trials for each disease (CT), the DALYs, and the API: $ln(CT_z) = \beta_0 + \beta_1 ln(API_z) + \beta_2 ln(DALYs_z) + e_z$. The estimated coefficients β_1 and β_2 are elasticities. Note that this model can have a "market size" interpretation, since it makes DALYs and API multiplicatively related to one another. Since we lack specific information about the supply function of clinical trials in each disease area, the error term will absorb idiosyncratic costs of drug development in each disease area. Recall that the model predicts that $\beta_2 = 0.5$.

Our chief results are shown in the first column of Table 2. That there is a larger response to API than to the health losses caused by a disease shows the importance of

ability to pay. The empirical results line up with the model's predicted value for the elasticity of the number of clinical trials with respect to DALYs.

VARIABLES	(1) Total Clinical Trials	(2) CTs with an industry partner	(3) CTs without any industry partner	(4) CTs for Cancer	(5) Total Clinical Trials
DALYs	0.542***	0.642***	0.503***	0.762***	0.532***
	(0.0702)	(0.0855)	(0.0681)	(0.128)	(0.0724)
API	0.886***	1.099***	0.786***	1.364***	0.985***
	(0.128)	(0.156)	(0.124)	(0.462)	(0.135)
# of drugs					0.132*
					(0.0705)
Constant	-10.88***	-15.25***	-10.04***	-18.48***	-12.08***
	(1.636)	(1.992)	(1.587)	(4.980)	(1.745)
Observations	94	94	94	26	88
R-squared	0.528	0.525	0.496	0.640	0.586

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2. Regression results showing the relationship between clinical trials, DALYs and Average Patient Income. API affected both industry and non-industry trials similarly.

It is unsurprising that when pharmaceutical companies are funding clinical trials, they do so with an eye towards profitability. However, 46% of the trials in our data list no industry sponsor or involvement. We separately tested trials with and without industry partners in Columns 2 and 3 of Table 2. For both industry and non-industry (governments, universities, foundations) trials, there is a large income effect, which is somewhat larger in the case of industry. If non-industry entities compensated for industry's preference for high-income diseases, the estimated coefficient on API in Column 3 would be negative.

Cancers constitute the only group of diseases large enough to test separately. Cancers, because they have some commonalities, are interesting to test by themselves;

their clinical trials are likely to be more similar in cost and success rates than clinical trials across all diseases. As Column 4 shows, even in this relatively homogeneous disease category, API remains important.

We also explored the effect of including in the regression the number of distinct approved drugs approved by the US FDA for each cause, where a drug was defined as an active pharmaceutical ingredient, matching by MeSH code from RxMix.("RxMix," n.d.) We then mapped them to each cause using the same correspondence as for clinical trials. The results are shown in Column 5 of Table 2. The effect of the number of drugs approved for each disease is relatively small and has weak statistical significance, which suggests that it is the *untreated* burden of disease that is important for the decision to invest in clinical trials of new treatments.

We also explored the relationship between the number of clinical trials and whether a disease was "neglected" or infectious. Column 1 of Table 3 shows that when accounting for only the total number of DALYs, there is a statistically significant negative coefficient on the Neglected Tropical Diseases ("NTD") dummy. However, as shown in Column 2, when we include API in the regression, the NTD effect becomes statistically insignificant. Thus, it appears that NTDs are neglected not because they are tropical, but because the patients are poor. Table 3 also allows us to explore whether, as has been claimed, non-communicable diseases are under-recognized (Horton, 2017). Column 3 shows that instead, infectious diseases tend to be under-researched relative to their global health importance: there is a large and significant negative coefficient on the Infectious dummy. However, as with NTDs, this effect appears to be driven by income, as shown in Column 4.

VARIABLES	(1) Total trials	(2) Total trials	(3) Total trials	(4) Total trials
DALYs	0.344***	0.512***	0.463***	0.550***
	(0.0888)	(0.0857)	(0.0795)	(0.0734)
API		0.823***		0.940***
		(0.163)		(0.194)
NTD	-1.968***	-0.323		
	(0.465)	(0.527)		
Infectious			-1.320***	0.152
			(0.299)	(0.405)
Constant	0.385	-9.822***	-1.383	-11.52***
	(1.407)	(2.381)	(1.251)	(2.376)
Observations	94	94	94	94
R-squared	0.398	0.530	0.406	0.529

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3. Regression results when controlling for NTD or infectious diseases.

One concern about the results observed in Tables 2 and 3 is whether the observed income effect might be due to different treatment levels across countries. If some diseases such as HIV/AIDS have adequate treatments available, the need for clinical trials should be reduced, since the reason for excessive DALYs for those diseases in low-income countries is likely a lack of *access* to treatment rather than inadequate existing treatments. This could result in biased results, as the diseases that are relatively poorly treated in low-income countries would have low API; at the same time, the value of investing in clinical trials might be small if existing treatments were effective but not being widely used. One way to get around this problem this is to use incidence, rather than DALYs, as the measure of health loss. Incidence, which is the rate of *new* cases in a population, should be only weakly related to how well a disease is treated, at least for non-communicable diseases. Column 1 of Table 4 shows a regression using incidence. API, when weighted by incidence rather

than by DALYs, has about the same coefficient as in Table 2. The coefficient on Incidence is quite low (though still statistically significant), which should not be surprising: there is more clinical research on diseases that cause more health loss, rather than diseases that are simply common. We replicated this regression using only non-communicable diseases, since their incidence is unlikely to be influenced by treatment. As Column 3 shows, the coefficient on API remains about the same, implying that differential treatment levels across countries is not driving the observed income effect. Column 2 shows the same regression as Column 1 except with DALYs rather than Incidence.

	(1)	(2)	(3)
VARIABLES	All Diseases	All Diseases	NCDs only
Incidence	0.129***		0.153**
	(0.0441)		(0.0598)
API (Incidence)	0.807***	0.818***	0.970***
	(0.170)	(0.131)	(0.336)
DALYs		0.548***	
		(0.0670)	
Constant	-3.729**	-10.41***	-5.538
	(1.821)	(1.674)	(3.764)
Observations	84	84	68
R-squared	0.250	0.546	0.130

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4. Regression results when using incidence rather than DALYs.

In the regressions shown elsewhere in this paper, we included clinical trials initiated during the years 2006-2015, and calculated DALYs averaged over 2005, 2010 and 2015, and income averaged over 2006-2015. We explored the effect of segmenting this data by time, so that trials during 2006-2010 are related to income and DALYs in 2005, while trials during 2011-2015 are related to income and DALYs in 2010. The results are

presented in columns 1 and 2 of Table 5, which are estimated using completely different data but have almost identical estimated coefficients for API. We explored regression using the first differences, but these did not generate statistically significant results, which is not surprising, given that changes in the burden of disease and income are slow and drug development is a long-term process.

VARIABLES	(1) Clinical Trials (2006-2010)	(2) Clinical Trials (2011-2015)
DALYs (2005)	0.516***	
	(0.0650)	
API (2005)	0.939***	
	(0.130)	
DALYs (2010)		0.538***
		(0.0694)
API (2010)		0.940***
		(0.132)
Constant	-11.57***	-11.96***
	(1.566)	(1.709)
Observations	94	93
R-squared	0.550	0.522

Table 5. Regression results with data segmented over time.

We performed numerous robustness checks to our main regression, with results shown in Appendix 2. Specifically, we considered alternative measures of health loss instead of DALYs. We explored the effect of separating clinical trials by phase. We segmented the data according to whether the disease had high or low API or DALYs. We segmented the data according to type of intervention – drug/biologic or non-drug therapy.

We added diseases for which there is an existing excellent therapy. In all cases, the estimated elasticities were consistent in sign and magnitude.

"Missing" investment

Governments and other institutions such as WHO often use disease characteristics to target research funding. Our analysis suggests that it would be useful to focus on diseases that disproportionately affect poor people. We expect that for-profit corporations will not target those diseases and conditions systematically, creating "missing" investment in relevant therapies.

If governments and other institutions were to fill in the investment in the development of new therapies for low-income diseases, what would be required? Using the estimated coefficient from the first column of Table 2, we simulated the number of Phase 1-3 clinical trials that would be expected for each disease, given its DALYs, if all had the average API. We calculated the income-normalized number of clinical trials $\widehat{CT}_i$, as $ln(\widehat{CT}_z) = ln(CT_z) + \hat{\beta}_2(ln(\overline{API}) - ln(API_z))$ where $\overline{API}$ is the average API for all 94 diseases and $\hat{\beta}_2$ is the estimated coefficient from the first column of Table 2. The averages of $\widehat{CT}_z$ and CT_z are slightly different. To be able to compare these two values, we rescaled $\widehat{CT}_z$ by multiplying it by $\overline{CT}/\overline{\widehat{CT}}$, where $\overline{CT}$ is the average of the actual number of clinical trials and $\overline{\widehat{CT}}$ is the average of the simulated number of clinical trials. This approach offers a data-driven way to think about where investment might be missing. For example, HIV actually had 92 trials per year: the income-normalized number of trials is 274. Ebola had 4 trials per year; the income-normalized number of trials is 58. We show the number of actual Phase 1-3 clinical trials and the income-normalized analogue in Appendix 1, along

with data on all trials, DALYs and API by disease. Evidently, this exercise is meant to be exploratory rather than dispositive, but we think it effectively illustrates the scale of the income effect.

We then applied standard measures of trial costs by phase to calculate the cost of bringing research in all diseases up to the current average. To calculate the total amount of missing investment, we first calculated the difference between actual and income-normalized clinical trials $\Delta CT_z = CT_z - \widehat{CT_z}$ for each disease for Phases 1-3. The average cost of clinical trials is assumed to be \$4m for phase 1, \$13m for phase 2, and \$20m for phase 3 (Sertkaya, Birkenbach, Berlind, & Eyraud, 2014). We multiplied the number of missing clinical trials ΔCT_z by the average clinical trial cost for that phase. We then calculated the sum *only* for diseases with below-average API and negative ΔCT_z, to obtain the required additional investment for each phase. In effect, we focus on diseases that appear to have little investment because of low API. As described above, the average number of diseases per clinical trial for phase 1 is 1.46, for phase 2 is 1.36 and for phase 3 is 1.46. To avoid double-counting, we divided the investment for each phase by the average number of diseases per clinical trial for each phase.

After applying cost per trial data, the incremental annual investment to bring all diseases up to the income-normalized level of Phase 1-3 clinical trials is approximately \$9bn, which compares to total estimated actual investment in clinical trials of \$44bn. This latter number is calculated by multiplying the number of clinical trials in each phase by the average cost per trial and adjusting for double-counting as above. (As a test of this estimate, we compared our estimated total clinical trials cost of \$44bn to commercial sources, which estimate current annual clinical trials expenditures of \$52.5bn, including

Phase 4 trials (INC research, 2016).) Trials depend on appropriate bench science, which costs roughly 75% of clinical trial expenses.(DiMasi et al., 2016) This implies that the investment required to eliminate severe underfunding of research into the diseases that disproportionately affect poor people is about $16bn annually, about 4 times the annual Gates Foundation grant funding. While this would not solve problems of access to new drugs, it is a snapshot of the investment necessary to bring *innovation* for low-income diseases up to the current average.

Discussion

This paper has developed a novel model of investment into health innovations, and then estimated it directly. We find a significant disparity in investment into treatments for diseases and conditions based on income. Specifically, the elasticity of the number of clinical trials with respect to API is close to 1; since calculated API ranges from approximately $500 to $30,000, this elasticity has significant implications for the allocation of research. Nevertheless, the estimated disparities are poorly characterized by the "10/90 Gap." It would require substantial investments to bring innovation in the low-income diseases up to the current average.

We have not attempted in this paper to present a normative argument for whether governments *ought* to invest more in innovation for diseases of the poor. We conclude the paper by considering this issue. It is typical in economics to use willingness to pay as a measure of value, leading to the conclusion that it is efficient to have low investment in the diseases that affect poor people, since their "willingness" to pay is low. In our view, this conclusion is unwarranted given the underlying theory, which does not support interpersonal comparisons of utility. Few people, including economists, would argue that

we should be indifferent between saving the life of a single person with an income of $25,000 and saving 200 people with an average income of $1,000. Yet this is the implication of the estimated elasticity of clinical trials with respect to API and DALYs.

Another common objection to investing in medical innovation for poor people is that poorly functioning health systems may block poor people from using treatments even if developed. However, this argument should not be overstated: first, our analysis shows that the income effect is present even if we consider only the upper half of diseases by API; and second, many medicines are effectively distributed even in very low-income countries. The demand for medicines is universal, although in some places it is not supported by much ability to pay. If R&D were underwritten by governments and foundations, then prices could be low and (in)ability to pay need not present an insuperable obstacle.

References

Acemoglu, D., & Linn, J. (2004). Market Size in Innovation: Theory and Evidence from the Pharmaceutical Industry. *The Quarterly Journal of Economics*, *119*(3), 1049–1090. https://doi.org/10.1162/0033553041502144

Barrenho, E., Miraldo, M., & Smith, P. C. (2019). Does global drug innovation correspond to burden of disease? The neglected diseases in developed and developing countries. *Health Economics (United Kingdom)*, *28*(1), 123–143. https://doi.org/10.1002/hec.3833

ClinicalTrials.gov. (2017). Protocol Registration Data Element Definitions for Interventional and Observational Studies. Retrieved August 10, 2017, from

https://prsinfo.clinicaltrials.gov/definitions.html#StudyPhase

Commission on Health Research for Development. (1990). *Health research: essential link to equity in development*. Oxford University Press, USA.

De Angelis, C., Drazen, J. M., Frizelle, F. A., Haug, C., Hoey, J., Horton, R., … Weyden, M. B. Van Der. (2004). Clinical Trial Registration: A Statement from the International Committee of Medical Journal Editors. *New England Journal of Medicine*, *351*(12), 1250–1251. https://doi.org/10.1056/NEJMe048225

DiMasi, J. A., Grabowski, H. G., & Hansen, R. W. (2016). Innovation in the pharmaceutical industry: New estimates of R&D costs. *Journal of Health Economics*, *47*, 20–33. https://doi.org/10.1016/j.jhealeco.2016.01.012

Dubois, P., de Mouzon, O., Scott-Morton, F., & Seabright, P. (2015). Market size and pharmaceutical innovation. *The RAND Journal of Economics*, *46*(4), 844–871. https://doi.org/10.1111/1756-2171.12113

Horton, R. (2017). Non-communicable diseases: 2015 to 2025. *The Lancet*, *381*(9866), 509–510. https://doi.org/10.1016/S0140-6736(13)60100-2

INC research. (2016). *Investor Presentation William Blair 36th Annual Growth Stock Conference*. Retrieved from http://www.investor.syneoshealth.com/static-files/d123b034-8ede-4ce1-a117-40a4d76760d9

Institute for Genome Sciences. (2019). Disease Ontology. Retrieved July 10, 2019, from http://disease-ontology.org/

Kyle, M. K., & McGahan, A. M. (2011). Investments in Pharmaceuticals before and after TRIPS. *Review of Economics and Statistics*, *94*(November), 1157–1172. https://doi.org/10.1162/REST_a_00214

Labonte, R., & Spiegel, J. (2003). Setting global health research priorities: Burden of disease and inherently global health issues should both be considered. *BMJ: British Medical Journal*, *326*(7392), 722. https://doi.org/10.1136/bmj.326.7392.722

Lichtenberg, F. R. (2007). Importation and Innovation. *Economics of Innovation and New Technology*, *16*(6), 403–417. https://doi.org/10.1080/10438590601002307

Lown, B., & Banerjee, A. (2006). The developing world in The New England Journal of Medicine. *Globalization and Health*, *2*, 3. https://doi.org/10.1186/1744-8603-2-3

Martino, O. I., Ward, D. J., Packer, C., Simpson, S., & Stevens, A. (2012). Innovation and the burden of disease: Retrospective observational study of new and emerging health technologies reported by the EuroScan network from 2000 to 2009. *Value in Health*, *15*(2), 376–380. https://doi.org/10.1016/j.jval.2011.11.034

Murray, C. J. L., Ezzati, M., Flaxman, A. D., Lim, S., Lozano, R., Michaud, C., … Lopez, A. D. (2012). GBD 2010: Design, definitions, and metrics. *The Lancet*. https://doi.org/10.1016/S0140-6736(12)61899-6

Perloff, J. M., & Salop, S. C. (1985). Equilibrium with product differentiation. *The Review of Economic Studies*, *52*(1), 107–120.

Ridley, D. B., Grabowski, H. G., & Moe, J. L. (2006). Developing Drugs For Developing Countries. *Health Affairs*, *25*(2), 313–324. https://doi.org/10.1377/hlthaff.25.2.313

RxMix. (n.d.). Retrieved April 19, 2018, from https://mor.nlm.nih.gov/RxMix/

Sertkaya, A., Birkenbach, A., Berlind, A., & Eyraud, J. (2014). *Examination of clinical trial costs and barriers for drug development*. *Report, US Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation, Washington, DC*.

UN General Assembly. (2012). Political Declaration of the High-level Meeting of the General

Assembly on the Prevention and Control of Non-communicable Diseases. A/RES/66/2. *UN New York*, *49777*(January), 1–13. https://doi.org/10.1007/BF03038934

US Food and Drug Administration. (2016). Guidance for Industry: Tropical Disease Priority Review Vouchers. Washington, DC.

Vidyasagar, D. (2006). Global notes: the 10/90 gap disparities in global health research. *Journal of Perinatology : Official Journal of the California Perinatal Association*, *26*(1), 55–56. https://doi.org/10.1038/sj.jp.7211402

Viergever, R. F., Terry, R. F., & Karam, G. (2013). Use of data from registered clinical trials to identify gaps in health research and development. *Bulletin of the World Health Organization*, *91*(6), 416–425. https://doi.org/10.2471/BLT.12.114454

Weber, W. E. J., Merino, J. G., & Loder, E. (2015). Trial registration 10 years on. *BMJ*, *351*(July), 2. https://doi.org/10.1136/bmj.h3572

World Bank. (2019). GNI per capita (constant 2010 US$).

World Health Organization. (2017). Neglected tropical diseases. Retrieved August 10, 2017, from http://www.who.int/neglected_diseases/diseases/en/